# Explaining the Value-Added Model

An Informal Guide to Understanding
Teacher and School Effectiveness Estimates in Plain English

Assessment, Research, and Data Analysis
Apr. 2012, version 1.5

*The background on new teacher evaluations*

On March 24, 2011 Governor Rick Scott signed into law the Student Success Act, a Florida law that changes the way the teachers in the state's public schools will be evaluated. Beginning with the 2011-12 school year, the Student Success Act calls for one-half of the teacher evaluation to be based on the ratings made by school administrators, and the other half to be based on the student achievement gains calculated by a complex statistical procedure known as the Value-Added Model (VAM). The intention of this paper is to provide a non-technical interpretation of the major outcomes of the VAM to assist in the understanding of these matters and inform future discussions.

*The need for VAM*

Students are not randomly assigned to teachers, and teachers are not randomly assigned to schools. These simple facts cause researchers some serious difficulties in teacher evaluation. Teachers start out with different kinds of classrooms -- different in beginning student achievement levels and different in student characteristics. Trying to adjust outcome measures to account for these initial differences is a very tricky business. The entire value-added methodology was constructed  to try to accommodate these complications. Simpler statistical methods are available, but do not treat the students as being "nested" within teachers, who are, in turn, nested within schools. The hierarchical nature of VAM allows for nesting, but with increased complexity comes increased difficulty in explaining and comprehending the technicalities.

*The importance of speaking clearly*

In elementary introductions of VAM statistics, casual wording of explanations may be acceptable, using whatever metaphors and over-simplifications might be effective in getting the basic ideas across. But now that most of us have acquired a passing acquaintance with the main features, and the numerical results are actually being used in teacher evaluations with real consequences, we need to be more careful with our language. It's important to try to talk in consistent terminology and speak in clear, unambiguous terms if we want to communicate our questions and understandings of these new statistical practices.

*Predictions of things that have already occurred*

Let's start with agreeing on the names of the outcomes. Each student, based on his or her previous performances and descriptive characteristics, has a *predicted score*. It's kind of odd that statisticians would refer to this as a prediction, since the real test results for the students were already available when the predictions were formulated. Usually, predictions are of future events, things we don't know yet. However, here we're not as interested in the predictions as we are in what's left of the student's score after we take out that which can be predicted from the nuisance variables. In this context, the statisticians are using the overall existing relationship between predictors and the known outcomes to calculate what would be the general trend for students' test scores based on their previous performance and their demographic descriptors. It would then be possible to compare each student's particular score to the overall trend to see if it is significantly different than the expectation. It would probably be more accurate to simply refer to these trends as *expected scores*, but we are somewhat forced to stick with the name *predicted scores* if for no other reason than everybody else will continue to use this label in this context.

*Student residuals and the percent of students exceeding expectations*

The difference between these predicted scores and the actual scores are the *student residuals* – positive differences if they exceed expectations, negative if they don't. The student residuals are the fundamental basis for the estimates of the teacher and school effects. They are the essential elements of the statistical mechanism that ultimately results in the official measures of teacher effectiveness. But we should take a moment here to appreciate that they can also be referred to explicitly by themselves as useful early-stage measures of success. The *percent of students exceeding expectations* is an appealing simple-to-compute, easy-to-understand concept. It can readily be aggregated at the teacher or the school level and may provide a straightforward, albeit less-sophisticated, indicator of effectiveness.

*Variables used in the prediction of student performance*

In the development of the predicted scores, a large number of student characteristics are considered, along with the student's test results from previous years. These student characteristics include such things as age, disability/gifted status, English language learner status, mobility, and attendance, as well as classroom variables, such as class size and homogeneity of previous student performance. The intention of including all of these variables into the statistical model is to "level the playing field" by accounting for differences in the proficiency and characteristics of students assigned to teachers. It should be noted that, because of adjustments made on the basis of these attribute variables, students in the same class with the same test performance history may, nonetheless, have different predicted scores.

*Variables **not** used in the prediction of student performance*

Somewhat obvious in their absence in the model are variables concerning race/ethnicity and socio-economic status. Although considered useful predictors in other research contexts, state law prohibits their use in teacher evaluations. Since these variables are known to be associated with other variables included in the model, it may be reasonable to assume that they are already being accounted for by student prior performance to some extent. Whether the extensive list of incorporated student variables has been sufficient in adjusting for differences in student assignment is an important question of endless debate among statisticians.

*Describing the teacher effects*

For each teacher, the differences between the expectations based on predicted scores and the students' actual performances are assessed. The unique *teacher effects* or *teacher components* are the portions of the achievement beyond expectations that the statistical procedure directly attributes to the individual teachers. They represent the "value-added" part of the student residual that is credited to the teacher's instruction and classroom environment. It would be nice if it was simply the average of the student residuals for a teacher, but it is somewhat more complicated in VAM. The calculated teacher effect is numerically expressed, at least in this early stage, in terms of the Development Scale Score (DSS) point advantage students would have by virtue of being in a particular teacher's class. Don't confuse this preliminary teacher effect with the final refined teacher VAM, to be discussed later.

*The School Component*

The *school component* refers to that part of the achievement beyond expectations that is common to all students in a school. It may reflect principal leadership, allocation of resources, school climate factors, as well as other local community influences. It, too, is expressed in DSS units and may be larger or smaller than the teacher effects in the school. Some people loosely refer to this measure as the school *effect*, but since it doesn't really cause any changes in student performance, the word *component* may be preferable.

*The Issue of Subtraction*

In earlier explanations of VAM, it was tempting to think of the school components as having been, at some time, subtracted from the teacher effects. "After all," the argument went, "if we never tried to account for school level components in the equations, all these effects would be attributed to the teachers." This conception, while perhaps helpful in framing the understanding of the origins of effects, is not entirely accurate. Schools were in the equations from the start and no subtracting, per se, was ever done. It is probably more appropriate to think of the VAM estimation process as the simultaneous partitioning of student achievement beyond expectations into that which can be associated with individual teachers and that which can be associated with schools.

*Common Teacher Effects*

There is yet a third kind of effect, not separately entered into the original statistical model, which we might refer to as the *common teacher effect*. It may be thought of as comprising teacher collaboration, support, and learning conditions as well as the average achievement advantage of all the teachers in a school. Because it is common to all teachers within a school, it naturally gets initially partitioned as part of the overall school component.

*Putting Part of the School Effects into the Teacher Effects*

The Student Growth Implementation Committee thought that the part of the school component that is really the common teacher effects should be "put back in" to the unique teacher effects to get a fair and complete estimate of teacher effectiveness. How much of the school component to put back in was a topic of much debate. In the absence of other compelling arguments, the committee compromised on 50 percent. The resultant *teacher VAM estimate* is the unique teacher effect with half of the school component added in. It is this teacher VAM estimate that is the numerical measure of effectiveness that is passed along to the next stage of constructing teacher evaluations.

*Normalizing Teacher Effects*

So far, we have been talking about things as if there were only one FCAT test. In fact, there are many – one for each grade level in each content area. Each of these tests has its own score characteristics and growth curves. Since the VAM estimates need to be combined and compared across grades, some kind of standardization is required. Although the State has offered a type of “normalized” teacher VAM estimate, it is the responsibility of each District to adopt a final normalization process. Each different standardization method would result in different VAM values and have an effect on the local interpretations of the results. In Miami-Dade, the teacher VAMs are normalized by the conventional method of converting to standardized scores with common averages and dispersions. The standardized scores allow for statistically appropriate methods of combining VAMs across grade levels and comparing VAMs between grad levels. Moreover, they can be conveniently interpreted in terms of percentile standing, which greatly simplifies the communication and understanding of the results. These teacher VAMs, expressed as percentile standings within grade level for the district, are final-level measures that would be used in teacher evaluations and other district statistical applications.